Nonlinear desirability as a linear classification problem

Arianna Casanova ARIANNA.CASANOVA@IDSIA.CH
Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), Switzerland

Alessio Benavoli ALESSIO.BENAVOLI@TCD.IE
School of Computer Science and Statistics, Trinity College Dublin, Ireland

Marco Zaffalon ZAFFALON@IDSIA.CH
Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), Switzerland

Abstract

The present paper proposes a generalization of linearity axioms of coherence through a geometrical approach, which leads to an alternative interpretation of desirability as a *classification problem*. In particular, we analyze different sets of rationality axioms and, for each one of them, we show that proving that a subject, who provides finite accept and reject statements, respects these axioms, corresponds to solving a binary classification task using, each time, a different (usually nonlinear) family of classifiers. Moreover, by borrowing ideas from machine learning, we show the possibility to define a *feature mapping* allowing us to reformulate the above nonlinear classification problems as linear ones in a higher-dimensional space. This allows us to interpret gambles directly as payoffs vectors of *monetary lotteries*, as well as to reduce the task of proving the rationality of a subject to a linear classification task.

Keywords: imprecise probabilities, coherence, convex coherence, monetary scale, piecewise separators

1. Introduction

The Bayesian framework is a sound and consistent theory because it is a *logic*. In fact, it can be shown that the rules of (Bayesian) probabilities can be inferred via mathematical duality from a set of *logical axioms* [3, 12, 23, 25, 30, 31], that one can interpret as rationality requirements in the way a subject accepts gambles on the results of an uncertain experiment. Mathematically, for a finite possibility space, gambles g are represented by column vectors in $\mathbb{R}^n$, where n is the size of the (finite) possibility space Ω ($n = |\Omega|$) of the experiment. In this paper, we interpret gambles as payoff vectors of *monetary* lotteries.[1] By denoting with $\mathscr{D}$ the set of gambles an agent finds to be desirable and with g a generic gamble in $\mathbb{R}^n$, the axioms of desirability can be expressed as:[2]

D1:	Tautologies	if $g \geq 0$ then $g \in \mathscr{D}$,
D2:	Falsum	if $g < 0$ then $g \notin \mathscr{D}$,
D3:	Linearity	if $g, h \in \mathscr{D}$ then $\lambda g + \mu h \in \mathscr{D}$ for any $\lambda, \mu \in \mathbb{R}^+$,
D4:	Closure	if $g + \varepsilon \in \mathscr{D}$ for all $\varepsilon \in \mathbb{R}^+_*$, then $g \in \mathscr{D}$,
D5:	Completeness	if $g \notin \mathscr{D}$ then $-g \in \mathscr{D}$.

The first axiom – D1 – also known as *accepting partial gains* criterion in literature [30], states that an agent should always accept non-negative gambles, because they can only increase the agent's utility. The second axiom – D2 –also known as *avoiding sure loss* criterion in literature [30], states that an agent should always avoid negative gambles, because they can only decrease the agent's utility. In what follows, for simplicity, we denote by T the set of all non-negative gambles $T := \{g \in \mathbb{R}^n : g \geq 0\}$ and by F the set of negative gambles $F := \{g \in \mathbb{R}^n : g < 0\}$. Hence, the two previous axioms can be rewritten respectively as D1: $T \subseteq \mathscr{D}$ and D2: $\mathscr{D} \cap F = \emptyset$. The third and fourth axiom state instead that the utility of the agent is linear – D3 (a standard assumption in decision theory) – and, respectively continuous: D4. D5 states that an agent has complete preferences on the states of the world.

These axioms provide the foundation of rational decision making as it was shown in [18, 32, 33], providing a connection between desirability and Von Neumann and Morgestern's [19] and Anscombe and Aumann's [3] axiomatisation of rationality. However, in several situations the above axioms can be restrictive and researchers have relaxed and generalised them in many ways. For instance, it is not very realistic to assume that an agent can always compare alternatives. Axiomatisations of rational decision making under *incompleteness* can be obtained by dropping D5 [4, 13, 14, 18, 29, 30]. The closure (continuity) axiom D4 can also be abandoned. Although D4 makes tighter the connection between desirability and probability theory [5, 8, 10], the extra generality of sets of desirable

1. This is not the usual interpretation. Walley [30] has discussed in some detail why gambles should be interpreted as uncertain rewards in terms of lottery tickets that can be won or lost. In this paper, we abandon this assumption because we aim to relax the linearity axiom.

2. In D4, we use ε to denote also the constant gamble given by $g(\omega) = \varepsilon$ for every $\omega \in \Omega$.

gambles without D4 makes them useful when dealing with the problem of conditioning on sets of probability zero, or of choosing between two options under zero expectation [28, 30]. The axioms D1 and D2 can also be restrictive. For instance, in case Ω is an infinite possibility space, evalutating the positivity (or negativity) of a gamble can be computationally very demanding. This leads to a notion of computational rationality [6, 7] which restricts D1 and D2. Instead [26] focuses on relaxations of the notion of avoiding sure loss (D2) only. Finally, the axiom D3 (linearity) is also not very realistic, especially when one considers monetary gambles (i.e. gambles that return euros instead of lottery tickets about euros). The linearity axiom has been replaced by convexity in [17, 20].

In this paper, we assume that D1, D2 and D4 hold true and provide a *geometric* view of desirability which allows us to easily extend desirability to nonlinear utility.

In order to introduce this interpretation, we first consider a slightly more general framework for desirability whereby an agent can express both acceptance or rejection of gambles [22]. By rejecting a gamble, the agent expresses that they consider accepting that gamble unreasonable. Let $\mathscr{A}$ and $\mathscr{R}$ denote the set of acceptable and, respectively, rejectable gambles by the agent. We assume then the presence of a *modeller*, who aims to prove the *rationality* of the agent. The modeller can have in mind different concepts of rationality depending on the axioms they decide to consider as *rational*. Once established the axioms, they should evaluate the rationality of the agent on the basis of the only information available, which we assume to be $(\mathscr{A},\mathscr{R})$. Clearly, in practical situations, $\mathscr{A}$ and $\mathscr{R}$ are finite. So, we assume that they consider the agent rational if there exists a set $\mathscr{D}$ such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D} \cap \mathscr{R} = \emptyset$ that satisfies the rationality axioms established by the modeller. In this paper, we assume that $\mathscr{D}$ must always satisfy D1, D2 and D4 (which we consider minimal requirements for rationality), therefore $T \subseteq \mathscr{D}$, $\mathscr{D} \cap F = \emptyset$ and $\mathscr{D}$ must be *closed* in the sense of D4. We will present rationality models where the modeller considers additional requirements of rationality and discuss methods to choose *the least committal* set $\mathscr{D}$ compatible with these requirements – $\mathscr{D}$ will be defined as *deductive closure* of $\mathscr{A}$ taken with respect to the set of rationality axioms considered.[3]

It is well-known that, if the modeller assumes all axioms of rationality D1$-$D5, then the problem of determining if the agent is rational or not becomes, thanks to the hyperplane separation theorem, a binary linear classification problem [30]. Indeed, in this case the agent is rational if and only if there exists a *linear prevision* $P(\cdot)$ such that $P(g) \geq 0$, for every $g \in \mathscr{A}$ and $P(g) < 0$, for every $g \in \mathscr{R}$. However, a linear prevision is essentially an expected value operator taken with respect to a finitely additive probability: $P(g) = g^T\beta$ for every $g \in \mathbb{R}^n$ with $\beta \in \mathbb{R}^n$, $\beta \gneq 0$ and $\sum_{i=1}^n \beta_i = 1$, where β_i, for every i, are the components of the vector β. This is equivalent to have a binary linear classifier that classifies any gamble g on the basis of the sign of $P(g) = g^T\beta$, which classifies in a different way $\mathscr{A}$ and $\mathscr{R}$. It is easy to see that the constraints on the coefficient β of this classifier, can be dropped asking it to directly classify in a different way $\mathscr{A} \cup T$ and $\mathscr{R} \cup F$.[4] In this case, we say that the pair $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ is *linearly separable*. Figure 1.1 provides a 2D illustration of this case.

In this paper, we show that this way of interpreting the problem as a binary classification task, is very general. In particular, indeed, we show that:

- in the *imprecise* case (that is, assuming D1–D4 but not *completness* axiom D5), the problem can be reformulated using a *binary piecewise linear classifier* (see Section 2);
- if rationality is expressed through D1, D2, convexity replacing linearity, and D4 [17], the problem can be reformulated using a *binary piecewise (convex) affine classifier* (see Section 3);
- if rationality is expressed through D1, D2, “positive additivity” replacing linearity, and D4, the problem can be reformulated using a *binary piecewise positive affine classifier* (see Section 4);
- for more general cases that, however, respect D1, D2 and D4, the problem can be reformulate using a *nonlinear classifier* (see Section 6).

This allows us to model more realistic cases of desirability, see Example 2. Moreover, by borrowing ideas from machine learning,[5] we show that we can define a *feature mapping* that allows us to reformulate the above nonlinear classification problems as linear classifiers in a higher-dimensional space.

2. Standard Imprecise probability

Suppose to have a modeller who considers as *rational* an agent who respects D1-D4. Furthermore suppose, as before, that the only information they possess consist of a finite set of gambles $\mathscr{A}$, the agent is willing to accept, and a finite set of gambles $\mathscr{R}$, they are willing to reject. Then we can assume that the modeller considers the agent rational if and only if there exists a set of gambles $\mathscr{D}$ that satisfies D1-D4 (that we call *coherent* from now on), such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D} \cap \mathscr{R} = \emptyset$.

Let us indicate with $\mathscr{L}(\Omega)$, or $\mathscr{L}$ when there is no possible ambiguity, the set of all gambles defined on a possibility

3. We assume it is determined by $\mathscr{A}$ and only indirectly by $\mathscr{R}$.

4. Indeed, if $g^T\beta \geq 0$ for every $g \in T$ and $g^T\beta < 0$ for every $g \in F$, then $\beta \gneq 0$. Therefore, it can be normalized obtaining a probability distribution on Ω.

5. There is however a substantial difference between the classification problem we aim to solving here and standard classification problems considered in machine learning. In our case, together with the training data set $(\mathscr{A},\mathscr{R})$, there are two infinite *sets* T,F which must also be classified. To our best knowledge, this problem has only be studied by Mangasarian and Wild [16] for support vector machines.

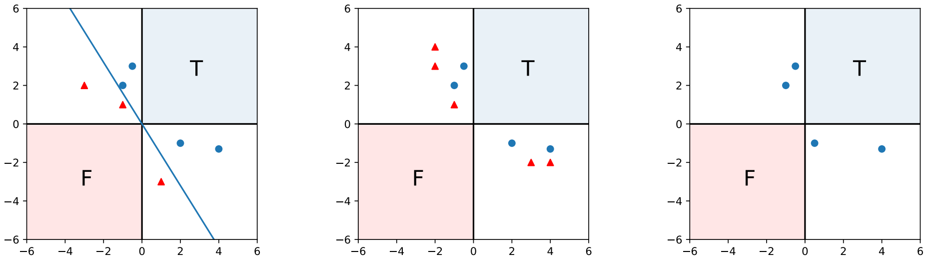

Figure 1.1: Example of a linearly separable pair $(\mathscr{A} \cup T, \mathscr{R} \cup F)$, (left). The middle and right plots instead, show two examples of *non-linearly separable* pairs $(\mathscr{A} \cup T, \mathscr{R} \cup F)$. The (blue) points denote gambles in $\mathscr{A}$; the (red) triangles denote gambles in $\mathscr{R}$. The blue line in the left plot is the linear function $g^T\beta = 0$ that separates $\mathscr{A} \cup T$ from $\mathscr{R} \cup T$. In the right plot, $\mathscr{R} = \emptyset$.

space Ω. Suppose moreover, as before, that $|\Omega| = n$. So, in what follows, $\mathscr{L} = \mathbb{R}^n$.

In this section we prove that, determining whether an agent is rational in the sense of respecting D1−D4, is equivalent to solve a *binary piecewise linear* classification task.

Definition 1 (Binary piecewise linear classifier) *We denote with the term* binary piecewise linear classifier, *a classifier $PLC(\cdot)$ defined on $\mathscr{L}$ and characterized by the following discriminant function:*[6]

$$PLC(g) := \begin{cases} 1 & \text{if } g^T\beta_j \geq 0, \text{ for all } j \in \{1,\dots N\} \\ -1 & \text{otherwise} \end{cases} \quad (1)$$

for every $g \in \mathscr{L}$, with $\beta_j \in \mathbb{R}^n$, for all j,[7] *$N \geq 1$.*

Definition 2 (Piecewise linear separability) *A pair of sets of gambles (A, B) is* piecewise linearly separable *if there exists a binary piecewise linear classifier $PLC(\cdot)$, such that $PLC(A) = 1$ and $PLC(B) = -1$.*[8] *In this case, we indicate the set of these classifiers with $PLC(A, B)$.*

Now we can show the main result of this section. All the proofs are in the supplementary material.

Proposition 3 *Given a pair of finite sets of gambles $(\mathscr{A}, \mathscr{R})$, there exists a coherent set $\mathscr{D}$, such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D} \cap \mathscr{R} = \emptyset$, if and only if $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ is piecewise linearly separable.*

From the proof of this proposition, it follows in particular that, if the pair $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ is piecewise linearly separable, the region classified as 1 by a classifier $PLC \in PLC(\mathscr{A} \cup T, \mathscr{R} \cup F)$, i.e. $\mathscr{D} := \{g \in \mathscr{L} : PLC(g) = 1\}$, is a coherent set, such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D} \cap \mathscr{R} = \emptyset$. Vice versa is not true in general.[9] Analogous reasonings can be applied for the other concepts of rationality, and hence of *coherence*, treated in Section 3 and Section 4. Obviously, the same is valid also for the most strictly concept of rationality (expressed through axioms D1-D5), seen in the Introduction (it is a subcase of the actual one).

In this case however, we can give another interpretation of a rational agent, which is an extension of the one seen in the Introduction. It is easy to show indeed that every binary piecewise linear classifier that classifies T as 1, is characterized by a set of weights $\{\beta_j\}_{j=1}^N$ such that $\beta_j \gneq 0$, for all $j = 1, \dots, N$. Therefore, we can normalise them getting N probability mass functions on Ω. So, given a pair of finite sets of gambles $(\mathscr{A}, \mathscr{R})$, $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ is piecewise linearly separable if and only if there exists a set of probability mass functions $P_j = \beta_j$ on Ω such that: $E_{P_j}(g)$, i.e. the expected value of a gamble g taken with respect to P_j, is not negative for all $g \in \mathscr{A} \cup T$ and all P_j; for all $g \in \mathscr{R} \cup F$ there exists at least a P_i such that $E_{P_i}(g) < 0$. Reformulation of Prop. 3 in those probabilistic terms is also a well-known result in the literature [30].

Once established that the agent is rational, i.e. if $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ is piecewise linearly separable, the minimal set of assumptions the modeller can made on the agent's beliefs is represented by the least committal coherent set containing $\mathscr{A}$ (see the proof of Prop. 3 or [30]):

$$\overline{\mathscr{E}(\mathscr{A})} := \overline{\mathsf{posi}(\mathscr{A} \cup T)}, \quad (2)$$

where, given $\mathscr{K} \subseteq \mathscr{L}$:

$$\mathsf{posi}(\mathscr{K}) := \left\{ \sum_{j=1}^{r} \lambda_j f_j : f_j \in \mathscr{K}, \lambda_j > 0, r \geq 1 \right\}, \quad (3)$$

and where $\overline{\mathscr{K}'}$ of a set $\mathscr{K}' \subseteq \mathscr{L}$ represents the closure of $\mathscr{K}'$ with respect to the supremum norm topology [30,

6. We take for simplicity the values 1 and −1.
7. W.l.o.g. we can assume $\beta_j \neq 0$ for every j since, otherwise, we can exclude it from the family of weights characterizing the classifier. This reasoning can be applied to every other classifier in the text.
8. With a little abuse of notation, with $PLC(\mathscr{K}) = c$ for $\mathscr{K} \subseteq \mathscr{L}$, $c \in \{-1, 1\}$, we mean $PLC(g) = c$, for all $g \in \mathscr{K}$. We will use the same notation also for the other types of binary classifiers considered later on.
9. For example if $\mathscr{D}$ is not *finitely generated* [30].

Section 3.7.2–3.7.4] or with respect to the usual topology of $\mathbb{R}^n$ [30, Appendix D]. Notice that, geometrically, it corresponds to a (closed) convex cone. Further, it corresponds also to a concept of *deductive closure* for the set $\mathscr{A}$. It is easy to see in fact that the operator $C_{cohr} : \mathscr{A} \to \overline{\mathscr{E}(\mathscr{A})}$, defined for every $\mathscr{A} \subseteq \mathscr{L}$, is a closure operator. The proof of Prop. 3 guarantees in particular that this set can be rewritten as the region classified as 1 by a binary piecewise inear classifier. Hence, in principle, it can be found solving the classification task with some other constraints, which can be formulated as a linear programming problem [9].

2.1. Feature Mapping

In this section we will prove that the previous problem, which is a non linear classification task in general, can be reformulated as a linear classification problem in a higher dimensional space. Let $\{\mathscr{B}_j\}_{j=1}^N$ denote a partition[10] of $\mathscr{L}$ defined as follows [21]:

$$\mathscr{B}_j := \{g \in \mathscr{L} : g^T \omega_j \leq g^T \omega_k \text{ for } k = 1, \dots, N, \quad j \neq k\}. \tag{4}$$

The vectors $\omega_j \in \mathbb{R}^n$ are parameters defining the partition. Now, we can introduce the feature mapping $\phi := (\phi_1, \dots, \phi_N)$, with $\phi_j : \mathbb{R}^n \to \mathbb{R}^n$ defined as $\phi_j : g \to \mathbb{I}_{\mathscr{B}_j}(g) g$, for every $j = 1, \dots, N$, where $\mathbb{I}_{\mathscr{B}_j}(g) := 1$ if $g \in \mathscr{B}_j$ and 0 otherwise.[11] Further, we define the following classifier, which corresponds to a linear classifier in the feature space:

$$LC_\phi(g) := \begin{cases} 1 & \text{if } \sum_{j=1}^N (\phi_j(g))^T \beta_j' \geq 0 \\ -1 & \text{otherwise.} \end{cases} \tag{5}$$

for every $g \in \mathscr{L}$, with $\beta_j' \in \mathbb{R}^n$, for all $j = 1, \dots, N$. In what follows, we consider both $\{\beta_j'\}_{j=1}^N$ and $\{\omega_j\}_{j=1}^N$ as parameters of $LC_\phi(\cdot)$. Finally, we introduce the following definition to simplify notation.

Definition 4 (Φ-separability) *A pair of sets of gambles (A, B) is Φ-separable, if there exists a classifier $LC_\phi(\cdot)$ of the type (5), such that $LC_\phi(A) = 1$ and $LC_\phi(B) = -1$. In this case, we indicate the set of these classifiers with $LC_\Phi(A, B)$.*

We can now state the main result of this section.

Proposition 5 *Let $(\mathscr{A}, \mathscr{R})$ be a pair of finite sets of gambles. If $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ is piecewise linearly separable, then it is also Φ-separable. Vice versa, if there exists a classifier $LC_\phi(\cdot) \in LC_\Phi(\mathscr{A} \cup T, \mathscr{R} \cup F)$ with $\omega_j = \beta_j'$, for all $j = 1, .., N$,[12] then $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ is also piecewise linearly separable.*

The proof of this proposition is based on the observation that: $\min(g^T\beta_1, \dots, g^T\beta_N) \geq 0$ if and only if $\sum_{j=1}^N (\mathbb{I}_{\mathscr{B}_j}(g) g)^T \beta_j \geq 0$, when $\omega_j = \beta_j$ for all j. Therefore, there is a one-to-one correspondence between classifiers $PLC(\cdot) \in \mathrm{PLC}(\mathscr{A} \cup T, \mathscr{R} \cup F)$ and classifiers $LC_\phi(\cdot) \in \mathrm{LC}_\Phi(\mathscr{A} \cup T, \mathscr{R} \cup F)$ with $\omega_j = \beta_j'$, for all $j = 1, .., N$. Further, in this case, the regions $\{g \in \mathscr{L} : LC_\phi(g) = 1\}$ correspond to coherent sets $\mathscr{D}$ such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D} \cap \mathscr{R} = \emptyset$. Moreover, $\beta_j' \gneq 0$, for all j, analogously to the constraints on β_j that we have for the correspondent classifiers $PLC(\cdot)$. Analogous considerations can be made for the feature mappings introduced for the other definitions of *coherence* treated in Section 3.1 and in Section 4.1.

Example 1 *Consider a coin tossing experiment whose outcomes are* h*, Heads and* t*, Tails. A gamble g in this case, has two components $g(\mathrm{h}) = g_1$ and $g(\mathrm{t}) = g_2$. If the agent accepts g then they commit themselves to receive/pay g_1 if the outcome is Heads and g_2 if Tails. Assume that $\mathscr{A} = \{[-1, 2]^T, [-0.5, 3]^T, [2, -1]^T\}$ and $\mathscr{R} = \{[-1, 1]^T, [-3, 2]^T, [1, -2]^T\}$, this could correspond to an imprecise rational agent whose $\mathscr{D} = \overline{\mathscr{E}(\mathscr{A})}$ is shown in Figure 2.1 (left). We can introduce the partition*

$$\mathscr{B}_1 = \{g : \tfrac{2}{6}g_1 + \tfrac{1}{6}g_2 \leq \tfrac{1}{6}g_1 + \tfrac{2}{6}g_2\}, \tag{6}$$

$$\mathscr{B}_2 = \{g : \tfrac{1}{6}g_1 + \tfrac{2}{6}g_2 \leq \tfrac{2}{6}g_1 + \tfrac{1}{6}g_2\}, \tag{7}$$

and verify that the classifier defined as:

$$LC_\phi(g) := \begin{cases} 1 & \text{if } \sum_{j=1}^2 (\mathbb{I}_{\mathscr{B}_j}(g) g)^T \beta_j' \geq 0 \\ -1 & \text{otherwise.} \end{cases} \tag{8}$$

for every $g \in \mathscr{L}$, with $\beta_1' = [\frac{2}{6}, \frac{1}{6}]^T$ and $\beta_2' = [\frac{1}{6}, \frac{2}{6}]^T$, classifies $\mathscr{A} \cup T$ as 1 and $\mathscr{R} \cup F$ as -1 and, moreover, the region $\{g \in \mathscr{L} : LC_\phi(g) = 1\}$ corresponds to $\overline{\mathscr{E}(\mathscr{A})}$.[13]

3. Convexity

When one considers monetary gambles, the linearity assumption D3 is not very realistic. In this section we consider a weaker form of rationality. We assume D1, D2, D4 and a relaxed version of the axiom D3:

10. We call it *partition* with a little abuse of notation. Indeed, we guarantee only that $\mathrm{int}\mathscr{B}_j \cap \mathrm{int}\mathscr{B}_k = \emptyset$, for every $j, k \in \{1, \dots, N\}$, $j \neq k$, where $\mathrm{int}\mathscr{B}_j$ represents the interior of $\mathscr{B}_j$in the usual topology of $\mathbb{R}^n$. Instead, it is guaranteed that $\cup_{j=1}^N \mathscr{B}_j = \mathscr{L}$ because every $g \in \mathscr{L}$ belongs to at least a $\mathscr{B}_j$. Indeed, for every $g \in \mathscr{L}$, $\{g^T\omega_j\}_{j \in N}$ is a finite set of real values, so the minimum always exists.

11. Therefore, $\mathbb{I}_{\mathscr{B}_j}(g)g = g$ if $g \in \mathscr{B}_j$ and 0_n otherwise, where 0_n is the null vector in $\mathbb{R}^n$. Analogous notation is used for the other feature mappings described in the article.

12. It is only a sufficient condition. Note also that the feature mapping defined in this section is not unique. We can find other feature mappings that guarantee the same properties (see Example 1).

13. Note that this is not the only feature mapping that we can use to linearly separate $\mathscr{A} \cup T$ from $\mathscr{R} \cup F$. Given the fact that $\mathscr{D}$ is a convex cone in 2D indeed, there always exist a linear classifier that classifies $\mathscr{A} \cup T$ as 1 and $\mathscr{R} \cup F$ as -1 in the feature space determined by the feature mapping $\eta : g \to [g_2, \mathbb{I}_{g_1<0}(g)g_1, \mathbb{I}_{g_1 \geq 0}(g)g_1]^T$. In this case, it is sufficient to consider $\beta' = [1, 2, \frac{1}{2}]^T$.

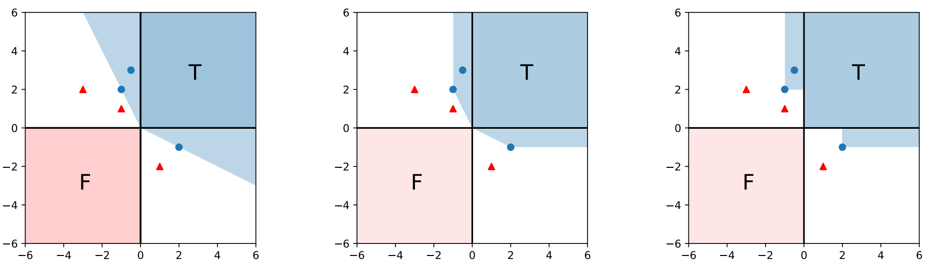

Figure 2.1: Example of a standard (imprecise) coherent set of gambles (left); a *convex* coherence set (center) and a *positive additive* coherent set (right). Points denote gambles in $\mathscr{A}$; triangles gambles in $\mathscr{R}$. The blue shaded region is the deductive closure of $\mathscr{A}$ for the three different notions of coherence.

D3* : if $g,h \in \mathscr{D}$, then $tg+(1-t)h \in \mathscr{D}$ for each $t \in [0,1]$ [Convexity].

We define a set $\mathscr{D}$ that satisfies D1, D2, D3*, D4 a *convex coherent* set of gambles.

Again, if an agent that is willing to accept and reject respectively a finite set $\mathscr{A}$ and a finite set $\mathscr{R}$, is *rational* in this sense, there exists a convex coherent set of gambles $\mathscr{D}$, such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D} \cap \mathscr{R} = \emptyset$. If this is the case, the minimal set of assumptions that the modeller can made on the agent's beliefs is represented by the least committal convex coherent set that contains $\mathscr{A}$. It can be proven that it corresponds to $\overline{\text{ch}(\mathscr{A} \cup T)}$, the *closed convex hull* of $\mathscr{A} \cup T$ in the usual topology of $\mathbb{R}^n$ or, equivalently, in the supremum norm topology (see Lemma 23). Furthermore, geometrically, it corresponds to a convex polyhedron (see Lemma 24). Clearly, also in this case, it corresponds to a deductive closure for the set $\mathscr{A}$ [15, Page 2039].

We claim that proving if an agent is rational in this sense is again equivalent to solve a binary classification task.

Definition 6 (Binary piecewise affine classifier) *We denote with the term* binary piecewise affine classifier *a classifier $PAC(\cdot)$ defined on $\mathscr{L}$ and characterized by the following discriminant function:*

$$PAC(g) := \begin{cases} 1 & \text{if } g^T\beta_j + \alpha_j \geq 0, \text{for all } j \in \{1,...N\} \\ -1 & \text{otherwise.} \end{cases} \tag{9}$$

for every $g \in \mathscr{L}$, with $\beta_j \in \mathbb{R}^n$, $\alpha_j \in \mathbb{R}$ for all j, $N \geq 1$.

Definition 7 (Piecewise affine separability) *A pair of sets of gambles (A,B) is* piecewise affine separable *if there exists a binary piecewise affine classifier $PAC(\cdot)$, such that $PAC(A) = 1$ and $PAC(B) = -1$. In this case, we indicate the set of these classifiers with $PAC(A,B)$.*

Now we can state the main result of this section.

Proposition 8 *Given a pair of finite sets of gambles $(\mathscr{A},\mathscr{R})$, there exists a convex coherent set of gambles $\mathscr{D}$, such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D} \cap \mathscr{R} = \emptyset$, if and only if $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ is piecewise affine separable.*

Analogously as before, from the proof of this proposition, it follows that if $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ is piecewise affine separable then $\mathscr{D} := \{g \in \mathscr{L} : PLC(g) = 1\}$ for every $PLC(\cdot) \in \text{PLC}(\mathscr{A} \cup T, \mathscr{R} \cup F)$, is a convex coherent set of gambles such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D} \cap \mathscr{R} = \emptyset$. Vice versa is not true in general, but it is true for $\overline{\text{ch}(\mathscr{A} \cup T)}$.

It is easy to show moreover that, again similarly as before, every binary piecewise affine classifier that classifies T as 1 and F as -1 is characterized by a set of weights $\{\beta_j, \alpha_j\}_{j=1}^N$ such that $\beta_j \gneq 0, \alpha_j \geq 0$, for all $j = 1,...,N$ with at least an $\alpha_k = 0$, $k \in \{1,\ldots,N\}$.

3.1. Feature Mapping

We can reformulate the previous problem as a linear classification task in a higher dimensional space, using a feature mapping similar to the one seen in Section 2.1. We can indeed define new partitions:

$$\begin{aligned} \mathscr{B}'_j := \{g' \in \mathscr{L}(\Omega') : {g'}^T \omega'_j \leq {g'}^T \omega'_k, \\ \text{for } k = 1,\ldots,N, \quad j \neq k\} \end{aligned} \tag{10}$$

with $|\Omega'| = n+1$, $\omega'_j \in \mathbb{R}^{n+1}$.

We can then introduce, the feature mapping $\psi := (\psi_1,\ldots,\psi_N)$, with $\psi_j : \mathbb{R}^n \to \mathbb{R}^{n+1}$, defined as $\psi_j : g \to I_{\mathscr{B}'_j}\left(\begin{bmatrix} g \\ 1 \end{bmatrix}\right)\begin{bmatrix} g \\ 1 \end{bmatrix}$ for every $j = 1,\ldots,N$. Further, we define the following classifier, which corresponds to a linear classifier in the feature space:

$$LC_\psi(g) := \begin{cases} 1 & \text{if } \sum_{j=1}^N (\psi_j(g))^T \beta'_j \geq 0 \\ -1 & \text{otherwise.} \end{cases} \tag{11}$$

for every $g \in \mathscr{L}$, with $\beta'_j \in \mathbb{R}^{n+1}$, for all $j = 1,\ldots,N$. We consider both $\{\beta'_j\}_{j=1}^N$ and $\{\omega'_j\}_{j=1}^N$ as parameters of $LC_\psi(\cdot)$. Then we can introduce the following definition.

Definition 9 (Ψ-separability) *A pair of sets of gambles (A,B) is* Ψ-separable, *if there exists a classifier $LC_\psi(\cdot)$ of the type (11), such that $LC_\psi(A) = 1$ and $LC_\psi(B) = -1$. In this case, we indicate the set of these classifiers with $LC_\Psi(A,B)$.*

We can now state the main result of this section.

Proposition 10 *Let $(\mathscr{A},\mathscr{R})$ be a pair of finite sets of gambles. If $(\mathscr{A}\cup T,\mathscr{R}\cup F)$ is piecewise affine separable, then it is also Ψ-separable. Vice versa, if there exists a classifier $LC_\psi(\cdot)\in LC_\Psi(\mathscr{A}\cup T,\mathscr{R}\cup F)$ with $\omega'_j=\beta'_j$, for all $j=1,..,N$, then $(\mathscr{A}\cup T,\mathscr{R}\cup F)$ is also piecewise affine separable.*

The proof of Prop. 10 is similar to that of Prop. 5, indeed it is based on an analogous observation. So, in particular, the region $\{g\in\mathscr{L}: LC_\psi(g)=1\}$ of a classifier $LC_\psi(\cdot)\in \mathrm{LC}_\Psi(\mathscr{A}\cup T,\mathscr{R}\cup F)$ with $\omega'_j=\beta'_j$ for all j, is, similarly as before, a convex coherent set $\mathscr{D}$ such that $\mathscr{D}\supseteq\mathscr{A}$ and $\mathscr{D}\cap\mathscr{R}=\emptyset$. Moreover, in this case, $\beta'_j\geq 0$, for all $j=1,\dots,N$, with $\beta'_{j,1:n}\gneq 0$, [14], with at least a β'_k, $k\in\{1,\dots,N\}$, such that $\beta'_{k,n+1}=0$.

Example 2 *Consider again Example 1. However now, suppose explicitly to work with* monetary *gambles, whose values represent rewards in* thousands of euros. *If the agent has limited financial resources, they can set their maximum loss to* 1 *thousand of euros, for example. In this case, it is no more reasonable thinking that they will respect axiom D*3. *A more reasonable rationality definition for the agent, is the one that is based on the axioms D*1*, D*2*, D*3**, D*4*. Analogously to Example 1, if we introduce a classifier $LC_\psi(\cdot)$ of the form (11), characterized by the parameters $\omega'_j=\beta'_j=\begin{bmatrix}\beta_j\\ \alpha_j\end{bmatrix}$ for $j=1,\dots,4$ where values of β_j,α_j for every j, correspond to the rows of the following matrices:*

$$\beta=\begin{bmatrix}1 & 0\\ \frac{2}{6} & \frac{1}{6}\\ \frac{1}{6} & \frac{2}{6}\\ 0 & 1\end{bmatrix},\quad \alpha=\begin{bmatrix}1\\0\\0\\1\end{bmatrix},$$

we can verify that $LC_\psi(\cdot)$ classifies $\mathscr{A}\cup T$ as 1 *and $\mathscr{R}\cup F$ as -1. Moreover, we have $\{g: LC_\psi(g)=1\}=\overline{\mathrm{ch}(\mathscr{A}\cup T)}$ (see proofs of Prop. 8 and 10). A graphical representation of $\overline{\mathrm{ch}(\mathscr{A}\cup T)}$ is showed in 2.1 (center).*

4. Positive additive coherence

We now consider an even weaker relaxation (see Lemma 22) of the linearity axioms D3:

D3** if $f\in\mathscr{D}$, then $f+h\in\mathscr{D}$ for each $h\in T$ [positive additivity].

We call a set $\mathscr{D}$ that satisfies D1, D2, D3**, D4 a *positive additive coherent* set of gambles. With a reasoning analogous to the previous sections, if, given a pair of finite sets $(\mathscr{A},\mathscr{R})$, there exists a positive additive coherent set of gambles $\mathscr{D}$, such that $\mathscr{D}\supseteq\mathscr{A}$ and $\mathscr{D}\cap\mathscr{R}=\emptyset$, the minimal such set is: $\uparrow(\mathscr{A}\cup\{0\})$ [11], as Lemma 25 proves. It is possible to notice that, geometrically, it corresponds to a union of orthants centered in the elements of $\mathscr{A}\cup\{0\}$. Further, also in this case, it is a deductive closure for $\mathscr{A}$ [2]. As usual, we show that proving if an agent is rational in this sense is equivalent to solve a binary classification task. Let us introduce the following definitions.

Definition 11 (PWP classifier) *We denote with the term* binary piecewise positive affine (PWP) classifier*, a classifier $PWPC(\cdot)$ defined on $\mathscr{L}$ and characterized by the following discriminant function:*

$$PWPC(g):=\begin{cases}1 & \text{if } \exists f^j\in\mathscr{F} \text{ s.t. } g\geq f^j,\\ -1 & \text{otherwise},\end{cases}\tag{12}$$

for every $g\in\mathscr{L}$, where $\mathscr{F}$ is a finite set of gambles.

Definition 12 (PWP separability) *A pair of sets of gambles (A,B) is* piecewise affine positive separable *(PWP) if there exists a PWP classifier $PWPC(\cdot)$, such that $PWPC(A)=1$ and $PWPC(B)=-1$. In this case we indicate the set of these classifiers with $PWPC(A,B)$.*

Note that, for every j, $\{g\geq f^j\}$ defines an orthant centered at f^j, whose border can be expressed as a piecewise affine function. It can easily be proved (by induction on the elements of $\mathscr{F}$) that the decision boundary of (12) is also a piecewise affine function. We can now state the main result of this section.

Proposition 13 *Given a pair of finite sets of gambles $(\mathscr{A},\mathscr{R})$, there exists a positive additive coherent set $\mathscr{D}$, such that $\mathscr{D}\supseteq\mathscr{A}$ and $\mathscr{D}\cap\mathscr{R}=\emptyset$, if and only if $(\mathscr{A}\cup T,\mathscr{R}\cup F)$ is PWP separable.*

Also here, if $(\mathscr{A}\cup T,\mathscr{R}\cup F)$ is PWP separable then $\mathscr{D}:=\{g\in\mathscr{L}: PWPC(g)=1\}$ for every $PWPC(\cdot)\in \mathrm{PWPC}(\mathscr{A}\cup T,\mathscr{R}\cup F)$, is a positive additive coherent set such that $\mathscr{D}\supseteq\mathscr{A}$ and $\mathscr{D}\cap\mathscr{R}=\emptyset$. Vice versa is not true in general, but it is true for $\uparrow(\mathscr{A}\cup\{0\})$.

4.1. Feature Mapping

In this section, we will prove that, also in this case, the previous problem can be reformulated as a linear classification task in a higher dimensional space. Let $\{\zeta_{i,j}\}$ with $i=1,..,n,\ j=1,..,N$ denote another partition of $\mathscr{L}=\mathbb{R}^n$ defined as follows:

$$\zeta_{i,j}:=\{g\in\mathscr{L}(\Omega):(g_i-\omega_i^j)=\max_k(\min_l(g_l-\omega_l^k))\}\tag{13}$$

with $\omega^j\in\mathbb{R}^n$, $j=1,...,N$. We can introduce the feature mapping $\rho:=(\rho_1,\dots,\rho_N)$, with $\rho_j:\mathbb{R}^n\to\mathbb{R}^{2n}$ defined as:

$$\rho_{j,i}(g):=\begin{cases}\mathbb{I}_{\zeta_{i,j}}(g)g_i & \text{if } i=1,...,n\\ \mathbb{I}_{\zeta_{i-n,j}}(g) & \text{otherwise}\end{cases}$$

14. $\beta'_{j,1:n}$ denotes the vector containing the first n components of β'_j.

for every $g \in \mathscr{L}$ and $j = 1,\dots,N$. Further we define the following classifier, which corresponds to a linear classifier in the feature space:

$$LC_\rho(g) := \begin{cases} 1 & \text{if } \sum_{j=1}^N (\rho_j(g))^T \beta'_j \geq 0 \\ -1 & \text{otherwise.} \end{cases} \tag{14}$$

for every $g \in \mathscr{L}$, with $\beta'_j \in \mathbb{R}^{2n}$. We consider both $\{\beta'_j\}_{j=1}^N$ and $\{\omega^j\}_{j=1}^N$ as parameters of $LC_\rho(\cdot)$. Similarly as before, we can introduce the following definition.

Definition 14 (P-separability) *A pair of sets of gambles (A,B) is P-separable, if there exists a classifier $LC_\rho(\cdot)$ of the type (14), such that $LC_\rho(A) = 1$ and $LC_\rho(B) = -1$. In this case, we indicate the set of these classifiers as $LC_P(A,B)$.*

We can now state the main result of this section.

Proposition 15 *Let $(\mathscr{A},\mathscr{R})$ be a pair of finite sets of gambles. If $(\mathscr{A}\cup T,\mathscr{R}\cup F)$ is PWP separable, then it is also P-separable. Vice versa, if there exists a classifier $LC_\rho(\cdot) \in LC_P(\mathscr{A}\cup T,\mathscr{R}\cup F)$ with $\beta'_{j,i} > 0$ and $\omega_i^j = -\frac{\beta'_{j,i+n}}{\beta'_{j,i}}$[15] for all i,j, then $(\mathscr{A}\cup T,\mathscr{R}\cup F)$ is also PWP separable.*

The proof is based on the following observation, analogous to the previous ones:

$$g \in \{PWPC(g) = 1\} \iff \max_j(\min_i(g_i - f_i^j)) \geq 0$$

$$\iff \sum_{j=1}^N \left(\begin{bmatrix} \mathbb{I}_{\zeta_{1,j}}(g) g_1 \\ \cdots \\ \mathbb{I}_{\zeta_{n,j}}(g) g_n \\ \mathbb{I}_{\zeta_{1,j}}(g) \\ \cdots \\ \mathbb{I}_{\zeta_{n,j}}(g) \end{bmatrix} \right)^T \begin{bmatrix} 1 \\ \cdots \\ 1 \\ -f_1^j \\ \cdots \\ -f_n^j \end{bmatrix} \geq 0.$$

if $\zeta_{i,j} = \{g \in \mathscr{L} : (g_i - f_i^j) = \max_k(\min_l(g_l - f_l^k))\}$, for all i,j, where f^j are the parameters of the PWP classifier $PWPC(\cdot)$. Also here the region $\{g \in \mathscr{L} : LC_\rho(g) = 1\}$ of a classifier $LC_\rho(\cdot) \in LC_P(\mathscr{A}\cup T,\mathscr{R}\cup F)$ with $\beta'_{j,i} > 0$ and $\omega_i^j = -\frac{\beta'_{j,i+n}}{\beta'_{j,i}}$ for every i,j, is, similarly as before, a positive additive coherent set $\mathscr{D}$ such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D} \cap \mathscr{R} = \emptyset$.

Example 3 *Consider again Example 1, but in this case we assume a positive additive rational agent, i.e. an agent who respects rationality axioms D1, D2, D3**, D4 (but not necessarily D3 or D3*). Analogously to Example 1 and 2, if we introduce a classifier $LC_\rho(\cdot)$ of the form (14), characterized by the parameters $\omega_i^j = -\beta'_{j,i+n}/\beta'_{j,i}$, and β'_j, $i = 1,2$, $j = 1,\dots,3$ defined in the following way:*

$$\beta'_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -2 \end{bmatrix}, \quad \beta'_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \beta'_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \\ 1 \end{bmatrix},$$

it classifies $\mathscr{A}\cup T$ as 1 and $\mathscr{R}\cup F$ as -1.[16] *Moreover, $\{g : LC_\rho(g) = 1\} = \uparrow(\mathscr{A}\cup\{0\})$ (see proofs of Prop. 13 and 15).*

5. Lower previsions and preferences

Consider a generic feature map $\chi = (\chi_1,\dots,\chi_N)$, with $\chi : \mathbb{R}^n \mapsto \mathbb{R}^M$ with $M \geq n$. From every classifier $LC_\chi(\cdot)$ defined as:

$$LC_\chi(g) := \begin{cases} 1 & \text{if } \sum_{j=1}^N (\chi_j(g))^T \beta'_j \geq 0 \\ -1 & \text{otherwise.} \end{cases} \tag{15}$$

for every $g \in \mathscr{L}$, with $\beta'_j \in \mathbb{R}^M$, for all $j = 1,\dots,N$ we can induce a preference relation on gambles of the original space $\mathbb{R}^n$ and therefore a lower and an upper prevision.

Definition 16 *Given a classifier $LC_\chi(\cdot)$ defined as in (15), we say that $f \succsim_{LC_\chi} g$ with $f,g \in \mathbb{R}^n$, if $LC_\chi(f-g) = 1$. In this case, for every gamble $g \in \mathbb{R}^n$, we call the following values:*

$$\underline{P}_{LC_\chi}(g) := \sup\{c \in \mathbb{R} : LC_\chi(g-c) = 1\},$$
$$\overline{P}_{LC_\chi}(g) := \inf\{c \in \mathbb{R} : LC_\chi(c-g) = 1\},$$

respectively the generalized lower *and the* generalized upper prevision *of g.*

Let us consider the case in which χ coincides with a feature mapping ϕ defined as in Section 2.1. Given a Φ-separable pair of sets $(\mathscr{A},\mathscr{R})$, from every classifier $LC_\phi(\cdot) \in LC_\Phi(\mathscr{A}\cup T,\mathscr{R}\cup F)$, we can induce a preference relation, and consequently a generalized lower and a upper prevision, on $\mathbb{R}^n$. In particular, the following result is valid.

Proposition 17 *Let $(\mathscr{A},\mathscr{R})$ be a pair of finite sets of gambles. If there exists a classifier $LC_\phi(\cdot) \in LC_\Phi(\mathscr{A}\cup T,\mathscr{R}\cup F)$ with $\omega_j = \beta'_j$, for all $j = 1,\dots,N$, then there exists a coherent set of gambles $\mathscr{D}$ (i.e. a set of gambles satisfying D1, D2, D3, D4), such that:*

$$f \succsim_{LC_\phi} g \iff f - g \in \mathscr{D}, \tag{16}$$

for every $f,g \in \mathscr{L}$. Moreover, if $f-g \in (\mathscr{A}\cup T)$, then $f \succsim_{LC_\phi} g$ and if $f-g \in (\mathscr{R}\cup F)$, then $(f \not\succsim_{LC_\phi} g)$. In particular, it follows that for every $g \in \mathscr{L}$:

$$\underline{P}_{LC_\phi}(g) = \underline{P}(g), \tag{17}$$

where $\underline{P}(g) := \sup\{c \in \mathbb{R} : g-c \in \mathscr{D}\}$ for every $g \in \mathscr{L}$, is the standard coherent lower prevision of g associated to $\mathscr{D}$.

The proof follows immediately from the proofs of Prop. 3 and 5. It follows then that the preference relation $\succsim_{LC_\phi}$, in this case, satisfies the following properties.

15. These constraints are not so restrictive. Indeed, it is computationally simple to verify if there exists such a classifier in this situation. It is sufficient to find a PWP classifier with $\mathscr{F} \subseteq (\mathscr{A}\cup\{0\})$ that classifies $\mathscr{A}\cup T$ as 1 and $\mathscr{R}\cup F$ as -1.

16. The coefficients β'_j of (12) can be determined by solving a simple mixed-integer linear programming problem.

R1. $-1 \not\succsim_{LC_\phi} 0$ [Avoiding Sure Loss];

R2. If $f \geq g$ then $f \succsim_{LC_\phi} g$, for any $f,g \in \mathscr{L}(\Omega)$ [Monotonicity];

R3. If $f \succsim_{LC_\phi} g$ and $\lambda > 0$ then $\lambda f \succsim_{LC_\phi} \lambda g$, for any $f,g \in \mathscr{L}(\Omega)$ [Positive homogeneity];

R4. If $f \succsim_{LC_\phi} g$ and $g \succsim_{LC_\phi} h$ then $f \succsim_{LC_\phi} h$, for any $f,g,h \in \mathscr{L}(\Omega)$ [Transitivity];

R5. If $f+\delta \succsim_{LC_\phi} g$ for every $\delta > 0$ then $f \succsim_{LC_\phi} g$, for any $f,g \in \mathscr{L}(\Omega)$ [Continuity];

R6. $f \succsim_{LC_\phi} g$ if and only if $f-g \succsim_{LC_\phi} 0$, for any $f,g \in \mathscr{L}(\Omega)$ [Cancellation].

We can prove similar results for the other cases.

Proposition 18 *Let $(\mathscr{A},\mathscr{R})$ be a pair of finite sets of gambles. If there exists a classifier $LC_\psi(\cdot) \in LC_\Psi(\mathscr{A} \cup T, \mathscr{R} \cup F)$ defined as in (11) with $\omega'_j = \beta'_j$, for all $j = 1,\dots,N$, then there exists a convex coherent set of gambles $\mathscr{D}$ (i.e. a set of gambles satisfying D1, D2, D3*, D4), such that:*

$$f \succsim_{LC_\psi} g \iff f-g \in \mathscr{D}, \tag{18}$$

for every $f,g \in \mathscr{L}$. Moreover, if $f-g \in (\mathscr{A} \cup T)$, then $f \succsim_{LC_\psi} g$ and if $f-g \in (\mathscr{R} \cup F)$, then $(f \not\succsim_{LC_\psi} g)$.

In this case, R1, R2, R5, R6 remains satisfied. Instead, R3 and R4 can be replaced by:

$$(f \succsim_{LC_\psi} g,\ h \succsim_{LC_\psi} l) \Rightarrow \alpha f + (1-\alpha)h \succsim_{LC_\psi} \alpha g + (1-\alpha) l$$
$$\text{for any } \alpha \in [0,1],\ f,g,h,l \in \mathscr{L}(\Omega),$$

that implies, in particular, the *convexity* axiom of [17].

Proposition 19 *Let $(\mathscr{A},\mathscr{R})$ be a pair of finite sets of gambles. If there exists a classifier $LC_\rho(\cdot) \in LC_P(\mathscr{A} \cup T, \mathscr{R} \cup F)$ with $\beta'_{j,i} > 0$ and $\omega^j_i = \frac{\beta'_{j,i+n}}{\beta'_{j,i}}$, for all $i = 1,\dots,n$, $j = 1,\dots,N$, then there exists a positive additive coherent set of gambles $\mathscr{D}$ (i.e. a set of gambles satisfying D1, D2, D3**, D4), such that:*

$$f \succsim_{LC_\rho} g \iff f-g \in \mathscr{D}, \tag{19}$$

for every $f,g \in \mathscr{L}$. Moreover, if $f-g \in (\mathscr{A} \cup T)$, then $f \succsim_{LC_\rho} g$ and if $f-g \in (\mathscr{R} \cup F)$, then $(f \not\succsim_{LC_\rho} g)$.

Also in this case, R1, R2, R5, R6 are still satisfied.

Example 4 *Consider again Example 1. As we have seen, the pair of sets $(\mathscr{A},\mathscr{R})$ is Φ-, Ψ- and P-separable. The three classifiers $LC_\phi(\cdot), LC_\psi(\cdot), LC_\rho(\cdot)$ considered in the previous examples moreover, respect the properties required by Props. 17, 18 and 19, respectively. In particular, they induce different preferences relations and different generalized lower previsions on gambles.*

In the following table are summarized the values of the three generalized lower previsions for $g = [2,-1]^T$, $g' = 0.5g$ and $g'' = 2g$. From these values, it is possible to notice that $\underline{P}_{LC_\psi}(\cdot)$ and $\underline{P}_{LC_\rho}(\cdot)$ are not coherent lower previsions (w.r.t. the standard coherence axioms for lower previsions). $\underline{P}_{LC_\phi}(\cdot)$ instead, coincides with a standard coherent lower prevision, as proven in Prop. 17.

	$\underline{P}_{LC_\phi}(\cdot)$	$\underline{P}_{LC_\psi}(\cdot)$	$\underline{P}_{LC_\rho}(\cdot)$
g	0	0	0
$g' = 0.5g$	0	0	-0.5
$g'' = 2g$	0	-1	-1

6. General nonlinear classifier

One may wonder what new kinds of applications are possible once we have redefined desirability as a classification problem. Here we provide the intuition for a general approach to define a nonlinear 'consequence operator' via a nonlinear classifier. This would allow us to perform inferences (computing lower previsions and preferences), from a pair of sets $(\mathscr{A},\mathscr{R})$, when we do not know the 'type of rationality' of the agent (i.e. we only assume D1, D2 and D4). To this purpose, we introduce a general feature mapping $\hat{\phi} = (\hat{\phi}_1,\dots,\hat{\phi}_N)$. Its j-th component $\hat{\phi}_j : \mathbb{R}^n \mapsto \mathbb{R}^M$ with $M \geq n$, is defined as:

$$\hat{\phi}_{ji}(g) := \begin{cases} \hat{\psi}_{ji}(g) g_i & \text{if } i \leq n, \\ \hat{\psi}_{ji}(g) & \text{otherwise.} \end{cases}$$

for every $g \in \mathscr{L}$. We assume that $\hat{\psi}_{ji}$ for every i,j, is a scalar function satisfying: (i) for every g, $\hat{\psi}_{ji}(g) \geq 0$, for all ji and exists kl with $l \leq n$, such that $\hat{\psi}_{kl}(g) > 0$; (ii) it depends on a vector of parameters $\theta \in \mathbb{R}^p$, $p \geq 1$, for all j,i. We then define the following classifier, which corresponds to a linear classifier in the feature space:

$$LC_{\hat{\phi}}(g) := \begin{cases} 1 & \text{if } \sum_{j=1}^N (\hat{\phi}_j(g))^T \beta'_j \geq 0 \\ -1 & \text{otherwise.} \end{cases} \tag{20}$$

As usual, we consider both $\{\beta'_j\}_{j=1}^N$ and θ as parameters of $LC_{\hat{\phi}}(\cdot)$. We introduce then the following definition.

Definition 20 ($\hat{\Phi}$-separability) *A pair of sets of gambles (A,B) is* $\hat{\Phi}$-separable *if there exist a classifier $LC_{\hat{\phi}}(\cdot)$ of the type (20), such that $LC_{\hat{\phi}}(A) = 1$ and $LC_{\hat{\phi}}(B) = -1$. In this case we indicate the set of these classifiers with $LC_{\hat{\Phi}}(A,B)$.*

It is easy to see that, depending on the form of $\hat{\psi}$, $\hat{\Phi}$-separability becomes Φ-separability, Ψ- separability or P-separability. However, in the previous sections we have seen that Φ, Ψ,P-separability of a pair $(\mathscr{A} \cup T, \mathscr{R} \cup F)$, where $\mathscr{A}$ and $\mathscr{R}$ are finite sets of gambles, is not sufficient to guarantee some form of rationality of the agent who is

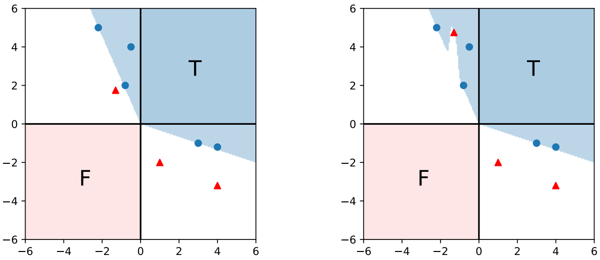

Figure 5.1: Nonlinear classifier for piecewise linearly separable (left) and for non piecewise linearly separable (right) $\mathscr{A},\mathscr{R}$.

willing to accept $\mathscr{A}$ and reject $\mathscr{R}$ (see Prop. 5, 10, 15). Nevertheless, it is possible to see that $\mathscr{D}=\{g : LC_{\hat{\phi}}(g)=1\}$, for every $LC_{\hat{\phi}}(\cdot)\in \mathrm{LC}_{\hat{\Phi}}(\mathscr{A}\cup T,\mathscr{R}\cup F)$ is such that $\mathscr{D}\supseteq\mathscr{A}$, $\mathscr{D}\cap\mathscr{R}=\emptyset$ and it satisfies D1, D2, D4.[17] Further, with this general family of feature mappings, we are potentially able to linearly separate in the feature space almost any pair $(\mathscr{A}\cup T,\mathscr{R}\cup F)$. So, it can be useful to constrain the possible feature mappings in such a way that the classifier, when its parameters are appropriately selected, is able to 'adapt' itself to the rationality of the subject. If, for example, $(\mathscr{A},\mathscr{R})$ is compatible with a stricter notion of coherence (for instance the pair $(\mathscr{A}\cup T,\mathscr{R}\cup F)$ is also piecewise linearly separable), then we would like that our classifier, when optimized, collapses to a classifier $LC_{\phi}(\cdot)$ defined in (5). We provide an example hereafter to clarify our intent.

Example 5 *Consider a possibility space of size $n=2$. Then, every gamble g has two components g_1 and g_2. Consider the feature mapping $\hat{\phi}(\cdot)$ whose non-null components are collected in the following vector:*

$$[g_2,\mathbb{I}_{g_1<0}(g)g_1,\mathbb{I}_{g_1\geq 0}(g)g_1,$$
$$\exp\left(\frac{-(g_1-r_1)^2}{\sigma_1^2}\right)g_1,\ldots,\exp\left(\frac{-(g_1-r_{|\mathscr{R}|})^2}{\sigma_{|\mathscr{R}|}^2}\right)g_1],$$

where r_i denotes the first component of the i-th gamble in $\mathscr{R}$ for $i=1,\ldots,|\mathscr{R}|$. There are $|\mathscr{R}|$ parameters, $\theta_i=\sigma_i^2$, which are the variances of the 'Gaussian' densities. We constrain $\beta'\gneq 0$ so that $LC_{\hat{\phi}}(T)=1$ and $LC_{\hat{\phi}}(F)=-1$. Among the classifiers that satisfy $LC_{\hat{\phi}}(\mathscr{A})=1$ and $LC_{\hat{\phi}}(\mathscr{R})=-1$, we select the one which minimises the objective function

$$\sum_{g_j\in\mathscr{A}}\left(\sum_{i=1}^{|\mathscr{R}|+3}\hat{\phi}_i(g_j)\beta_i'\right)+\gamma\sum_{i=4}^{|\mathscr{R}|+3}\beta_i', \tag{21}$$

with $\gamma\geq 0$. Note that $\sum_{i=1}^{|\mathscr{R}|+3}\hat{\phi}_i(g_j)\beta_i'\geq 0$ for $g_j\in\mathscr{A}$ (coherence constraint). Therefore, by minimizing the first term in the objective function, we are minimizing the sum of the distances[18] between the points $g_j\in\mathscr{A}$ and the hyperplane $\sum_{i=1}^{|\mathscr{R}|+3}\hat{\phi}_i(g)\beta_i'=0$. This means we are looking for the smallest extension in the feature space: the smallest set compatible with the assessments $\mathscr{A}$. The second term penalises the coefficients β_i' of the terms $\hat{\phi}_i(\cdot)$ involving the gaussian densities. This forces the classifier to be piecewise linear when $(\mathscr{A}\cup T,\mathscr{R}\cup F)$ are piecewise linearly separable. The above nonlinear nonconvex optimisation problem can be solved numerically. We provide two examples using $\gamma=100$ and $\mathscr{A}=\{(-0.5,4),(-2.2,5),(-0.8,2.0),(3,-1),(4,-1.2)\}$. Figure 5.1 (left) shows a case where $\mathscr{R}=\{(-1.3,1.75),(1,-2),(4,-3.2)\}$ and so $(\mathscr{A}\cup T,\mathscr{R}\cup F)$ are piecewise linearly separable. The solution of (21) gives $\beta_1'=1,\beta_2'=2.272,\beta_3'=0.33$, $\beta_4'=\beta_5'=\beta_6'=0$. In Figure 5.1 (right) $\mathscr{R}=\{(-1.3,4.75),(1,-2),(4,-3.2)\}$ and so $(\mathscr{A}\cup T,\mathscr{R}\cup F)$ are not piecewise linearly separable. The solution of (21) gives $\beta_1'=1,\beta_2'=2.272,\beta_3'=0.33$, $\beta_4'=3.45,\beta_5'=\beta_6'=0$ and $\sigma_1^2=0.038$. The lower prevision of the gamble $(-0.5,4)$ is 0.87 in the first case, and 0.54 in the second case.

7. Conclusions

We provided a new interpretation of (nonlinear) desirability as a classification problem. We have considered three instances of nonlinear desirability – imprecision, convex coherence and positive additive coherence – and showed that they can be expressed within this general framework. There are several research directions we aim to pursue in future works: (i) minimum coherence desiderata for the feature mapping; (ii) marginalisation and conditioning. Besides, given the connection between desirability and Von Neumann-Morgenstern's axiomatization of rationality [18, 32, 33], we plan to investigate if this general framework allows us to represent non-expected utility theory.

17. Thanks to the assumptions required on $\hat{\psi}$, the problem boils down to classifying only $\mathscr{A}$ and $\mathscr{R}$ providing the coefficients β_j' satisfy some constraints similar to the ones seen in the previous sections.

18. Technically, the distance is $(\sum_{i=1}^{|\mathscr{R}|+3}\hat{\phi}_i(g)\beta_i')/||\beta'||$, but w.l.o.g. we can rescale $\beta_i'/||\beta'||$.

References

[1] Chapter 2: A tutorial on polyhedral convex cones. In Rick Greer, editor, *Trees and Hills: Methodology for Maximizing Functions of Systems of Linear Relations*, volume 96 of *North-Holland Mathematics Studies*, pages 15 – 81. North-Holland, 1984.

[2] A. M. Abdalla. *Binary Closure Operators*. PhD thesis, Stellenbosch: Stellenbosch University, 2016.

[3] F. J. Anscombe, R. J. Aumann, et al. A definition of subjective probability. *Annals of Mathematical Statistics*, 34(1):199–205, 1963.

[4] R. J. Aumann. Utility theory without the completeness axiom. *Econometrica: Journal of the Econometric Society*, pages 445–462, 1962.

[5] A. Benavoli, A. Facchini, J. Vicente-Perez, and M. Zaffalon. A polarity theory for sets of desirable gambles. In *Proc. ISIPTA'17 Int. Symposium on Imprecise Probability: Theories and Applications,*, pages 1–12. PMLR, 2017.

[6] A. Benavoli, A. Facchini, D. Piga, and M. Zaffalon. Sum-of-squares for bounded rationality. *International Journal of Approximate Reasoning*, 105:130 – 152, 2019.

[7] A. Benavoli, A. Facchini, and M. Zaffalon. Computational complexity and the nature of quantum mechanics. *arxiv 1902.04569*, 2019.

[8] L. Blume, A. Brandenburger, and E. Dekel. Lexicographic probabilities and choice under uncertainty. *Econometrica: Journal of the Econometric Society*, pages 61–79, 1991.

[9] Inés Couso and Serafín Moral. Sets of desirable gambles: conditioning, representation, and precise probabilities. *International Journal of Approximate Reasoning*, 52(7):1034–1055, 2011.

[10] F. G. Cozman. Some remarks on sets of lexicographic probabilities and sets of desirable gambles. In *9th ISIPTA*, Pescara, Italy, 2015.

[11] B. A. Davey and H. A Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 2002.

[12] B. de Finetti. La prévision: ses lois logiques, ses sources subjectives. In *Annales de l'institut Henri Poincaré*, volume 7, pages 1–68, 1937.

[13] T. Galaabaatar and E. Karni. Subjective expected utility with incomplete preferences. *Econometrica*, 81(1):255–284, 2013.

[14] F. J. Girón and S. Ríos. Quasi-bayesian behaviour: A more realistic approach to decision making? *Trabajos de Estadistica Y de Investigacion Operativa*, 31(1): 17, 1980.

[15] C. Kiselman. A semigroup of operators in convexity theory. *Transactions of the American Mathematical Society*, 354(5):2035–2053, 2002.

[16] O. L. Mangasarian and E. W. Wild. Nonlinear knowledge-based classification. *IEEE Transactions on Neural Networks*, 19(10):1826–1832, 2008.

[17] R. Nau. Indeterminate probabilities on finite sets. *The Annals of Statistics*, pages 1737–1767, 1992.

[18] R. Nau et al. The shape of incomplete preferences. *The Annals of Statistics*, 34(5):2430–2448, 2006.

[19] L. J. Neumann, O. Morgenstern, et al. *Theory of Games and Economic Behavior*, volume 60. Princeton university press Princeton, 1947.

[20] R. Pelessoni and P. Vicig. Uncertainty modelling and conditioning with convex imprecise previsions. *International Journal of Approximate Reasoning*, 39 (2-3):297–319, 2005.

[21] D. Piga, A. Bemporad, and A. Benavoli. Rao-Blackwellized sampling for batch and recursive Bayesian inference of Piecewise Affine models. *Automatica*, 117, 2020.

[22] E. Quaeghebeur, G. de Cooman, and F. Hermans. Accept & reject statement-based uncertainty models. *International Journal of Approximate Reasoning*, 57: 69–102, 2015.

[23] F. P. Ramsey. Truth and probability. In *In R. B. Braithwaite, editor, The Foundations of Mathematics and other Logical Essays, chapter 7*, pages 156–198. Brace & Co., 1931.

[24] R. T. Rockafellar. *Convex Analysis*, volume 36. Princeton university press, 1970.

[25] L. J. Savage. *The Foundations of Statistics*. Wiley, New York, 1954.

[26] M. J. Schervish, T. Seidenfeld, and J. B. Kadane. How sets of coherent probabilities may serve as models for degrees of incoherence. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 8(03):347–355, 2000.

[27] A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1998.

[28] T. Seidenfeld, M. J. Schervish, and J. B. Kadane. Decisions without ordering. In *Acting and reflecting*, pages 143–170. Springer, 1990.

[29] T. Seidenfeld, M. J. Schervish, J. B. Kadane, et al. A representation of partially ordered preferences. *The Annals of Statistics*, 23(6):2168–2217, 1995.

[30] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1991.

[31] P. M. Williams. Notes on conditional previsions. Technical report, University of Sussex, 1975.

[32] M. Zaffalon and E. Miranda. Axiomatising incomplete preferences through sets of desirable gambles. *Journal of Artificial Intelligence Research*, 60:1057–1126, 2017.

[33] M. Zaffalon and E. Miranda. Desirability foundations of robust rational decision making. *Synthese*, pages 1–42, 2018.

Appendix A. Proofs of the main results

Proposition 21 *Consider a set of gambles $\mathscr{D} \subseteq \mathscr{L}$.*

If it is closed under the supremum norm topology, then it satisfies D4. Vice versa, if $\mathscr{D}$ satisfies also the following property:

$$f \geq g,\ g \in \mathscr{D} \Rightarrow f \in \mathscr{D} \tag{22}$$

then D4 implies closure in the supremum norm topology.

Proof It is well-known that $\mathscr{L}$ is a Banach space under the supremum norm and it is a linear topological space (with finite dimension n in our case) under the topology generated by the supremum norm (see [30]).

Now, consider $\mathscr{D}$ closed under the supremum norm topology. Then, the limit of every convergent sequence $(f_n)_{\{n\in\mathbb{N}\}}$ (respect to the supremum norm) with $f_n \in \mathscr{D}$ for every n, must be contained in $\mathscr{D}$. Consider then, a gamble f such that $f+\delta \in \mathscr{D}$ for every $\delta > 0$, then $f+\frac{1}{n} \in \mathscr{D}$ for every $n \in \mathbb{N}^*$. Its limit w.r.t. the supremum norm is f and, from the closure of $\mathscr{D}$, we know that $f \in \mathscr{D}$.

On the other hand, suppose $\mathscr{D}$ satisfies D4 and (22). Let us consider a succession $(f_n)_{\{n\in\mathbb{N}\}} \in \mathscr{D}$ convergent w.r.t. the sumpremum norm to a gamble $f \in \mathscr{L}$. We know that for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $\sup|f_n - f| < \varepsilon$ for all $n \geq N$. In particular, this means that there exist $h \in \mathscr{L}$ such that:

$$f_n - f = h^+ - h^-,\ \sup|h| < \varepsilon \tag{23}$$

hence:

$$f = (f_n + h^-) - h^+ \tag{24}$$

but, $f_n + h^- \in \mathscr{D}$ by hypothesis, and $f = (f_n+h^-) - h^+ \geq (f_n+h^-) - \varepsilon$. Then $f+\varepsilon \geq (f_n+h^-) \in \mathscr{D}$, from which it follows that $f+\varepsilon \in \mathscr{D}$. This procedure can be repeated for every $\varepsilon > 0$. Then by D4, we have $f \in \mathscr{D}$. ■

Proof [Proof of Proposition 3] Consider a pair of finite sets $(\mathscr{A}, \mathscr{R})$ for which there exists a coherent set of gambles $\mathscr{D}$, such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D} \cap \mathscr{R} = \emptyset$. Then, the minimal coherent set $\mathscr{D}$ that satisfies these conditions is $\overline{\mathscr{E}(\mathscr{A})} := \overline{\mathsf{posi}(\mathscr{A} \cup T)}$, where $\mathsf{posi}(\mathscr{K}) := \left\{\sum_{j=1}^r \lambda_j f_j : f_j \in \mathscr{K}, \lambda_j > 0, r \geq 1\right\}$ for every $\mathscr{K} \subseteq \mathscr{L}(\Omega)$ and where $\overline{\mathscr{K}'}$ of a set $\mathscr{K}' \subseteq \mathscr{L}$ represents the closure of $\mathscr{K}'$ with respect to the supremum norm topology. In fact, $\mathscr{E}(\mathscr{A})$ is clearly the minimal set $\mathscr{D}$ that satisfies D1 - D3 such that $\mathscr{D} \supseteq \mathscr{A}$. Then, thanks to Proposition 21, $\overline{\mathscr{E}(\mathscr{A})}$ is the minimal coherent set $\mathscr{D}'$ such that $\mathscr{D}' \supseteq \mathscr{A}$ and clearly, by hypothesis, we know also that $\overline{\mathscr{E}(\mathscr{A})} \cap \mathscr{R} = \emptyset$. This fact is also well-known in literature [30].

However, $\overline{\mathscr{E}(\mathscr{A})}$, by definition, is a polyhedral (convex) cone [1, Definition 2.3.2]. Indeed $\overline{\mathscr{E}(\mathscr{A})}$ can be rewritten as:

$$\overline{\mathscr{E}(\mathscr{A})} = \overline{\mathsf{posi}(\mathscr{A} \cup T)} =$$

$$C := \left\{g : g = \sum_{j=1}^r \lambda_j f_j,\ f_j \in (\mathscr{A} \cup \{\mathbb{I}_{\omega_i}\}_{i=1}^n), r \geq 1, \lambda_j \geq 0\right\}$$

where the last equality derives from the facts that: $\mathscr{E}(\mathscr{A}) = \mathsf{posi}(\mathscr{A} \cup T)$ is generated by the finite set $(\mathscr{A} \cup \{\mathbb{I}_{\omega_i}\}_{i=1}^n)$; C is already closed under the usual topology of $\mathbb{R}^n$ that coincides with the closure with respect to the supremum norm topology, for every topological space with n dimension [30, Appendix D]. The latter is true because, thanks to the Minkowsky-Weyl theorem [1], we know that C is an intersection of a finite number of closed halfspaces whose bounding hyperspaces pass through the origin:

$$C = \{g : g^T \beta_j \geq 0,\ j = 1, \ldots, N\} \tag{25}$$

with $\beta_j \in \mathbb{R}^n$. This concludes this part of the proof since it tells us that there exists a binary piecewise linear classifier $PLC(\cdot)$ with parameters β_j, which classifies $\mathscr{A} \cup T \subseteq \overline{\mathscr{E}(\mathscr{A})} = C =: \{g \in \mathscr{L} : PLC(g) = 1\}$ as 1 and $(\mathscr{R} \cup F)$, that has empty intersection with C, as -1.

Vice versa, consider a piecewise linearly separable pair $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ and a classifier $PLC(\cdot) \in \mathrm{PLC}(\mathscr{A} \cup T, \mathscr{R} \cup F)$. Then:

$$\{g : PLC(g) = 1\} = \{g : g^T \beta_j \geq 0, \text{ for all } j = 1, .., N\} \tag{26}$$

for some $\beta_j \in \mathbb{R}^n$ such that $\beta_{ji} \geq 0$, $\sum_i \beta_{ji} = 1$, for all i, j (constraints on β_j easily follow from the fact that $PLC(\cdot)$ classifies T as 1). Hence there exists a linear prevision P_j, such that $P_j(g) = g^T \beta_j$, for all g, for all $j = 1, \ldots, N$ [30, Section 2.8,Section 3.2]. Therefore we have:

$$\{g : PLC(g) = 1\} =$$
$$\{g : P_j(g) \geq 0, \text{for all} j = 1, .., N\} = \{g : \underline{P}(g) \geq 0\},$$

where $\underline{P} := \min_j\{P_j\}$ is a coherent lower prevision [30, Theorem 3.3.3]. Hence, $\mathscr{D} := \{g : PLC(g) = 1\}$ is a coherent set of gambles [30, Theorem 3.8.1].

In particular, we have also that $\mathscr{A} \subseteq \{g : PLC(g) = 1\} = \mathscr{D}$ and $\mathscr{R} \cap (\{g : PLC(g) = 1\} = \mathscr{D}) = \emptyset$ by hypotheses. ■

Proof [Proof of Proposition 5] Consider a piecewise linearly separable pair $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ and a classifier $PLC(\cdot) \in \mathrm{PLC}(\mathscr{A} \cup T, \mathscr{R} \cup F)$ with parameters $\{\beta_j\}_{j=1}^N$.

Then, a classifier $LC_\phi(\cdot)$ of the type (5) with parameters $\omega_j = \beta_j$ and $\beta'_j = \beta_j$ for all $j = 1, \ldots, N$, classifies $\mathscr{A} \cup T$ as 1 and $\mathscr{R} \cup F$ as -1. Indeed, consider $g \in \mathscr{L}$ and let us define $m := \min(g^T \beta_1, \ldots, g^T \beta_N)$. Then:

$$\sum_{j=1}^N (\phi_j(g))^T \beta_j = \sum_{j=1}^N (\mathbb{I}_{\mathscr{B}_j}(g) g)^T \beta_j = \sum_{k=1}^K g^T \beta_k = Km,$$

where, for every j, $\mathscr{B}_j$ are the partitions of the type 4 with $\omega_j = \beta_j$ and $g^T \beta_k = m$, for all $k = 1, \ldots, K$, with $1 \leq K \leq N$.

Hence, g is classified in the same way by the classifiers $PLC(\cdot)$ and $LC_\phi(\cdot)$. Therefore, in particular, if $g \in (\mathscr{A} \cup T)$, $m \geq 0$ and hence $\sum_{j=1}^{N}(\phi_j(g))^T\beta_j = Km \geq 0$, if instead $g \in (\mathscr{R} \cup F)$ then $m < 0$ and hence $\sum_{j=1}^{N}(\phi_j(g))^T\beta_j = Km < 0$.

Vice versa, let us consider a Φ-separable pair $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ and let us suppose the existence of a classifier $LC_\phi(\cdot) \in \mathrm{LC}_\Phi(\mathscr{A} \cup T, \mathscr{R} \cup F)$ with parameters $\omega_j = \beta'_j$, for all $j = 1, ..., N$. Let us define $m' := \min(g^T\beta'_1, ..., g^T\beta'_N)$. Then, for any $g \in \mathscr{L}$ we have:

$$\sum_{j=1}^{N}(\phi_j(g))^T\beta'_j = \sum_{k=1}^{K} g^T\beta'_k = Km', \tag{27}$$

where again $g^T\beta'_k = m'$, for all $k = 1, \ldots K$, with $1 \leq K \leq N$. Let us consider a binary piecewise linear classifier $PLC(\cdot)$ with parameters $\{\beta'_j\}_{j=1}^{N}$. Then, again, g is classified in the same way by the classifiers $LC_\phi(\cdot)$ and $PLC(\cdot)$. This is in particular true for $g \in \mathscr{A} \cup T$ and $g \in \mathscr{R} \cup F$. This means also that $\beta'_j \gneq 0$, for all $j = 1, .., N$. ■

Lemma 22 *If a set $\mathscr{D} \subseteq \mathscr{L}$, satisfies D1, D3* and D4 then it satisfies (22).*

Proof Consider $f \geq g$ with $g \in \mathscr{D}$. Then $f = g + t$ with $t \in T$. For any $\varepsilon > 0$, $f + \varepsilon = g + t + \varepsilon$. Moreover, we can always find $\lambda \in (0,1)$ such that $\lambda g \leq g + \varepsilon$.

Therefore, we have $f + \varepsilon = \lambda g + (1-\lambda)\frac{(g+\varepsilon-\lambda g)+t}{1-\lambda}$. Now, $g \in \mathscr{D}$ by hypothesis and $\frac{(g+\varepsilon-\lambda g)+t}{1-\lambda} \in T$, so $f + \varepsilon \in \mathscr{D}$. This can be repeated for every $\varepsilon > 0$, then $f + \varepsilon \in \mathscr{D}$ for all $\varepsilon > 0$ that implies, by D4, that $f \in \mathscr{D}$. ■

Lemma 23 *Given a pair of finite sets $(\mathscr{A}, \mathscr{R})$ for which there exists a convex coherent set of gambles $\mathscr{D}$ such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D} \cap \mathscr{R} = \emptyset$, then the minimal such set is $\mathscr{D} = \overline{\mathrm{ch}(\mathscr{A} \cup T)}$.*

Proof $\overline{\mathrm{ch}(\mathscr{A} \cup T)}$ satisfies D1 by definition and D3* [24, Theorem 6.2] and D4, thanks to Proposition 21.

Let us indicate with $\mathrm{D}(\mathscr{A}, \mathscr{R})$, the class of convex coherent sets of gambles $\mathscr{D}$ such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D} \cap \mathscr{R} = \emptyset$. Thanks to Lemma 22 and Proposition 21, every $\mathscr{D} \in \mathrm{D}(\mathscr{A}, \mathscr{R})$, is a convex closed set (respect to the topology of $\mathbb{R}^n$ or equivalently respect to the supremum norm topology) that contains $(\mathscr{A} \cup T)$.

Given the fact that $\overline{\mathrm{ch}(\mathscr{A} \cup T)} \supseteq \mathscr{A} \cup T$ and, by definition, it is the intersection of all the closed (respect to the topology of $\mathbb{R}^n$ or equivalently respect to the supremum norm topology) and convex sets containing $(\mathscr{A} \cup T)$, we have that $\overline{\mathrm{ch}(\mathscr{A} \cup T)} \subseteq \mathscr{D}$, for all $\mathscr{D} \in \mathrm{D}(\mathscr{A}, \mathscr{R})$.

But, every $\mathscr{D} \in \mathrm{D}(\mathscr{A}, \mathscr{R})$, satisfies $\mathscr{D} \cap (\mathscr{R} \cup F) = \emptyset$. Therefore, $\overline{\mathrm{ch}(\mathscr{A} \cup T)} \cap (\mathscr{R} \cup F) = \emptyset$, and hence it is also the smallest set $\mathscr{D} \in \mathrm{D}(\mathscr{A}, \mathscr{R})$. This concludes the proof. ■

Lemma 24 *Consider a finite set $\mathscr{A} \subseteq \mathscr{L}$. Then:*

$$\overline{\mathrm{ch}(\mathscr{A} \cup T)} = \mathrm{ch}^+(\mathscr{A} \cup \{0\}) := \{g : g \geq f,\ f \in \mathrm{ch}(\mathscr{A} \cup \{0\})\}.$$

Proof First of all, we can observe that:

$$\mathrm{ch}^+(\mathscr{A} \cup \{0\}) = \{g : g \geq f,\ f \in \mathrm{ch}(\mathscr{A} \cup \{0\}) =$$
$$= \sum_{i \in I} \alpha_i g_i + \sum_{j \in J} \gamma_j e_j =: \mathrm{ch}(\mathscr{A} \cup \{0\}) + \mathrm{posi}(e_1, \ldots, e_n)$$

with I, J finite, $g_i \in \mathscr{A} \cup \{0\}$, $\alpha_i, \gamma_i \geq 0$ and $\sum_i \alpha_i = 1$, where e_i is the canonical basis in $\mathbb{R}^n$ and $\mathrm{posi}(e_1, \ldots, e_n)$ is a convex polyhedral cone. From [27, Corollary 7.1.b], it follows that $\mathrm{ch}^+(\mathscr{A} \cup \{0\})$ is a convex (closed) polyhedron. Hence $\overline{\mathrm{ch}^+(\mathscr{A} \cup \{0\})} = \mathrm{ch}^+(\mathscr{A} \cup \{0\})$. Now, we divide the proof in two parts.

- $\overline{\mathrm{ch}(\mathscr{A} \cup T)} \subseteq \mathrm{ch}^+(\mathscr{A} \cup \{0\})$. Notice that, thanks to the previous observation, it is sufficient to show that $\mathrm{ch}(\mathscr{A} \cup T) \subseteq \mathrm{ch}^+(\mathscr{A} \cup \{0\})$. So, let us consider $g \in \mathrm{ch}(\mathscr{A} \cup T)$. By definition, we have:

$$g = \sum_{k=1}^{r} \lambda_k g_k$$

with $\lambda_k \geq 0$, for all $k = 1, ..., r$, $r \geq 1$, $\sum_{k=1}^{r} \lambda_k = 1$, $g_k \in (\mathscr{A} \cup T)$. Let us indicate with $Ind_{A \setminus T} := \{k \in \{1, ..., r\}$ such that : $g_k \in \mathscr{A} \setminus T\}$ and $Ind_T := \{k \in \{1, ..., r\}$ such that : $g_k \in T\}$. Then we have:

$$g \geq \sum_{k \in Ind_{A \setminus T}} \lambda_k g_k + \sum_{k \in Ind_T} \lambda_k 0,$$

hence $g \in \mathrm{ch}^+(\mathscr{A} \cup \{0\})$.

- $\mathrm{ch}^+(\mathscr{A} \cup \{0\}) \subseteq \overline{\mathrm{ch}(\mathscr{A} \cup T)}$. By definition, $\overline{\mathrm{ch}(\mathscr{A} \cup T)}$ is a closed convex set that contains T. Therefore, from Proposition 21 and Lemma 22, we have:

$$\mathrm{ch}(\mathscr{A} \cup \{0\}) \subseteq \overline{\mathrm{ch}(\mathscr{A} \cup T)} \Rightarrow$$
$$\mathrm{ch}^+(\mathscr{A} \cup \{0\}) \subseteq \overline{\mathrm{ch}(\mathscr{A} \cup T)}.$$

■

Proof [Proof of Proposition 8] Consider a pair of sets $(\mathscr{A}, \mathscr{R})$ for which there exists a convex coherent set of gambles $\mathscr{D}$, such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D} \cap \mathscr{R} = \emptyset$. Then the minimal convex coherent set $\mathscr{D}$, which satisfies these conditions, is $\overline{\mathrm{ch}(\mathscr{A} \cup T)}$. Thanks to Lemma 24, we know that it can be rewritten as:

$$\overline{\mathrm{ch}(\mathscr{A} \cup T)} = \mathrm{ch}^+(\mathscr{A} \cup \{0\}), \tag{28}$$

where $\mathrm{ch}^+(\mathscr{A} \cup \{0\})$ is a convex polyhedron. Any convex polyhedron can be written as an intersection of hyperspaces, whose border is a piecewise affine function. Therefore, there exists a piecewise affine classifier $PAC(\cdot)$, such that $\overline{\mathrm{ch}(\mathscr{A}\cup T)} = \mathrm{ch}^+(\mathscr{A}\cup\{0\}) = \{g : PAC(g) = 1\}$. Note moreover that $\overline{\mathrm{ch}(\mathscr{A}\cup T)} = \{g : PAC(g) = 1\} \supseteq (\mathscr{A}\cup T)$ and $(\overline{\mathrm{ch}(\mathscr{A}\cup T)} = \{g : PAC(g) = 1\}) \cap (\mathscr{R}\cup F) = \emptyset$ by construction.

Vice versa, consider a piecewise affine separable pair $(\mathscr{A}\cup T, \mathscr{R}\cup F)$. Let us consider a piecewise affine classifier $PAC(\cdot) \in \mathrm{PAC}(\mathscr{A}\cup T, \mathscr{R}\cup F)$. Now, the set:

$$\begin{aligned}\mathscr{D} :=&\{g : PAC(g) = 1\} = \\ &\{g : g^T\beta_j + \alpha_j \geq 0, \text{ for all } j = 1,...,N\}\end{aligned}$$

for some $\beta_j \in \mathbb{R}^n$ with $\beta_j \gneq 0$ and $\alpha_j \in \mathbb{R}$ for all $j \in \{1,\dots N\}$, is a convex coherent set of gambles such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D}\cap\mathscr{R} = \emptyset$. Indeed:

- $T \subseteq \mathscr{D}$ and $\mathscr{D}\cap F = \emptyset$, by definition, hence it satisfies D1 and D2;
- $\mathscr{D}$ satisfies D3*. Consider $g_1, g_2 \in \mathscr{D}$. Then $tg_1 + (1-t)g_2 \in \mathscr{D}$, for all $t \in [0,1]$. Indeed,
$$\begin{aligned}&(tg_1 + (1-t)g_2)^T\beta_j + \alpha_j = \\ &(tg_1)^T\beta_j + ((1-t)g_2)^T\beta_j + t\alpha_j + (1-t)\alpha_j = \\ &t((g_1)^T\beta_j + \alpha_j) + (1-t)(g_2^T\beta_j + \alpha_j) \geq 0\end{aligned}$$
for all $j \in \{1,...,N\}$.
- $\mathscr{D}$ is closed in the usual topology of $\mathbb{R}^n$ because it is the intersection of a finite number of closed half-spaces hence, thanks to Proposition 21, it satisfies D4.

Clearly, by the fact that $PAC(\cdot) \in \mathrm{PAC}(\mathscr{A}\cup T, \mathscr{R}\cup F)$, it is also true that $\mathscr{A} \subseteq \mathscr{D}$ and $\mathscr{D}\cap\mathscr{R} = \emptyset$. ■

Proof [Proof of Proposition 10]

Consider a piecewise affine separable pair $(\mathscr{A}\cup T, \mathscr{R}\cup F)$ and a classifier $PAC(\cdot) \in \mathrm{PAC}(\mathscr{A}\cup T, \mathscr{R}\cup F)$ with parameters $\{\beta_j, \alpha_j\}_{j=1}^N$.

Then, a classifier $LC_\Psi(\cdot)$ of the type (11) with parameters $\omega_j' = \beta_j' = \begin{bmatrix}\beta_j\\ \alpha_j\end{bmatrix}$, for all $j = 1,...,N$, classifies $\mathscr{A}\cup T$ as 1 and $\mathscr{R}\cup F$ as -1. Indeed, consider $g \in \mathscr{L}$ and let us define $m := \min(g^T\beta_1 + \alpha_1, \dots, g^T\beta_N + \alpha_N)$. Then:

$$\sum_{j=1}^N (\psi_j(g))^T \begin{bmatrix}\beta_j\\ \alpha_j\end{bmatrix} = \sum_{j=1}^N \left(\mathbb{I}_{\mathscr{B}'_j}\left(\begin{bmatrix}g\\ 1\end{bmatrix}\right)\begin{bmatrix}g\\ 1\end{bmatrix}\right)^T \begin{bmatrix}\beta_j\\ \alpha_j\end{bmatrix} =$$

$$\sum_{k=1}^K (g^T\beta_k + \alpha_k) = Km,$$

where, for every j, $\mathscr{B}'_j$ are the partitions of the type 10 with $\omega_j' = \beta_j$ and $g^T\beta_k + \alpha_k = m$, for any $k = 1,...,K$, with $1 \leq K \leq N$. Hence, g is classified in the same way by the classifiers $PAC(\cdot)$ and $LC_\Psi(\cdot)$. Therefore, in particular, if $g \in (\mathscr{A}\cup T)$, $m \geq 0$ and hence $\sum_{j=1}^N (\psi_j(g))^T \begin{bmatrix}\beta_j\\ \alpha_j\end{bmatrix} = Km \geq 0$, if instead $g \in (\mathscr{R}\cup F)$ then $m < 0$ and hence $\sum_{j=1}^N (\psi_j(g))^T \begin{bmatrix}\beta_j\\ \alpha_j\end{bmatrix} < 0$.

Vice versa, let us consider a Ψ-separable pair $(\mathscr{A}\cup T, \mathscr{R}\cup F)$ and let us suppose the existence of a classifier $LC_\Psi(\cdot) \in \mathrm{LC}_\Psi(\mathscr{A}\cup T, \mathscr{R}\cup F)$ with parameters $\omega_j' = \beta_j'$, for all $j = 1,...,N$. Let us define $m' := \min(g^T\beta'_{1,1:n} + \beta'_{1,n+1}, ..., g^T\beta'_{N,1:n} + \beta'_{N,n+1})$. Then, for any $g \in \mathscr{L}$, we have:

$$\sum_{j=1}^N (\psi_j(g))^T\beta_j' = \sum_{k=1}^K (g^T\beta'_{k,1:n} + \beta'_{k,n+1}) = Km',$$

where $\beta'_{k,1:n}$ is the vector containing the first n components of β'_k, for every k, and where again $(g^T\beta'_k + \beta'_{k,n+1}) = m'$, for all $k = 1,\dots K$, with $1 \leq K \leq N$. Let us consider a binary piecewise affine classifier $PAC(\cdot)$ with parameters $\{\beta'_{j,1:n}, \beta'_{j,n+1}\}_{j=1}^N$. Then, again, g is classified in the same way by the classifiers $LC_\Psi(\cdot)$ and $PAC(\cdot)$. This is in particular true for $g \in \mathscr{A}\cup T$ and $g \in \mathscr{R}\cup F$. This means also that $\beta'_{j,1:n} \gneq 0$, for all $j = 1,..,N$ and $\beta'_{j,n+1} \geq 0$, for all $j = 1,..,N$, with at least a $\beta'_{k,n+1} = 0$. ■

Lemma 25 *Given a pair of finite sets $(\mathscr{A},\mathscr{R})$ for which there exists a positive additive coherent set of gambles $\mathscr{D}$, such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D}\cap\mathscr{R} = \emptyset$, then the smallest such set is:*

$$\mathscr{D} = \uparrow(\mathscr{A}\cup\{0\}) := \{g : (\exists f \in \mathscr{A}\cup\{0\})\ g \geq f\}.$$

Proof $\uparrow(\mathscr{A}\cup\{0\})$ satisfies D1, D3** and $\mathscr{A} \subseteq \uparrow(\mathscr{A}\cup\{0\})$ by construction. Moreover, it satisfies also D4 by Proposition 21, because it is closed respect to the usual topology of $\mathbb{R}^n$ (it is a finite union of closed sets).

Let us indicate with $\mathrm{P}(\mathscr{A},\mathscr{R})$, the class of positive additive coherent sets of gambles $\mathscr{D}$, such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D}\cap\mathscr{R} = \emptyset$. Clearly, each $\mathscr{D} \in \mathrm{P}(\mathscr{A},\mathscr{R})$ satisfies $\mathscr{D} \supseteq \uparrow(\mathscr{A}\cup\{0\})$. But, every $\mathscr{D} \in \mathrm{P}(\mathscr{A},\mathscr{R})$, satisfies also $\mathscr{D}\cap(\mathscr{R}\cup F) = \emptyset$. Therefore, $\uparrow(\mathscr{A}\cup\{0\})\cap(\mathscr{R}\cup F) = \emptyset$. So, it is also the smallest positive additive coherent set of gambles $\mathscr{D} \in \mathrm{P}(\mathscr{A},\mathscr{R})$. ■

Proof [Proof of Proposition 13] Consider a pair of sets $(\mathscr{A},\mathscr{R})$ for which there exists a positive additive coherent set of gambles $\mathscr{D}$, such that $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D}\cap\mathscr{R} = \emptyset$. Then the minimal such set is $\uparrow(\mathscr{A}\cup\{0\})$. However, it can be rewritten as:

$$\uparrow(\mathscr{A}\cup\{0\}) = \{g \in \mathscr{L} : PWPC(g) = 1\}$$

where $PWPC(\cdot)$ is a PWP classifier, defined as:

$$PWPC(g) := \begin{cases}1 & \text{if } \exists f^j \in (\mathscr{A}\cup\{0\}) \text{ s.t. } g \geq f^j, \\ -1 & \text{otherwise.}\end{cases}$$

Therefore, given that $\mathscr{A}\cup T \subseteq \uparrow(\mathscr{A}\cup\{0\}) = \{g : PWPC(g) = 1\}$ and $(\uparrow(\mathscr{A}\cup\{0\}) = \{g : PWPC(g) =$

$1\}) \cap (\mathscr{R} \cup F) = \emptyset$, we have that $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ is PWP separable. Vice versa, consider a PWP separable pair $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ and a classifier $PWPC(\cdot) \in \mathrm{PWPC}(\mathscr{A} \cup T, \mathscr{R} \cup F)$. Then:

$$\mathscr{D} := \{g : PWPC(g) = 1\}$$

is, by construction, a positive additive coherent set of gambles. Indeed, it clearly satisfies D1, D2, D3**. Further, it is closed because it is a finite intersection of closed sets (respect to the usual topology of $\mathbb{R}^n$) hence, by Proposition 21, it satisfies D4. It satisfies also $\mathscr{D} \supseteq \mathscr{A}$ and $\mathscr{D} \cap \mathscr{R} = \emptyset$ by hypothesis. ■

Proof [Proof of Proposition 15] Consider a PWP separable pair $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ and a classifier $PWPC(\cdot) \in \mathrm{PWPC}(\mathscr{A} \cup T, \mathscr{R} \cup F)$ with parameters $\mathscr{F} = \{f^j\}_{j=1}^N$.

Then, a classifier $LC_\rho(\cdot)$ of the type (14), with parameters $\omega^j = f^j$ and $\beta'_j = \begin{bmatrix} 1 \\ \cdots \\ 1 \\ -f_1^j \\ \cdots \\ -f_n^j \end{bmatrix}$, for all $j = 1, \dots, N$, classifies $\mathscr{A} \cup T$ as 1 and $\mathscr{R} \cup F$ as -1.

Indeed,consider $g \in \mathscr{L}$ and let us define $m := \max_k(\min_l(g_l - f_l^k))$. Then:

$$\sum_{j=1}^N (\rho_j(g))^T \beta'_j = \sum_{j=1}^N \sum_{i=1}^n \mathbb{I}_{\zeta_{ij}}(g)(g_i - f_i^j) = KLm$$

where, for every i, j, ζ_{ij} are the partitions of the type 13 with $\omega^j = f^j$ and where $1 \le L \le n$, $1 \le K \le N$. Hence, g is classified in the same way by the classifiers $PWPC(\cdot)$ and $LC_\rho(\cdot)$. Therefore, in particular, if $g \in (\mathscr{A} \cup T)$, $m \ge 0$ and hence $LC_\rho(g) = 1$. If instead $g \in (\mathscr{R} \cup F)$ then $m < 0$ and hence $LC_\rho(g) = -1$.

Vice versa, let us consider a P-separable pair $(\mathscr{A} \cup T, \mathscr{R} \cup F)$ and let us suppose the existence of a classifier $LC_\rho(\cdot) \in \mathrm{LC}_P(\mathscr{A} \cup T, \mathscr{R} \cup F)$ with parameters $\{\beta'_j\}_{j=1}^N$ such that $\beta'_{j,i} > 0$, and $\omega_i^j = -\frac{\beta'_{j,i+n}}{\beta'_{j,i}}$ for all $i = 1, .., n$, $j = 1, ..., N$. Let us define $m' := \max_k(\min_l(g_l - (-\frac{\beta'_{k,l+n}}{\beta'_{k,l}})))$. Then, for any $g \in \mathscr{L}$:

$$\sum_{j=1}^N (\rho_j(g))^T \beta'_j = \sum_{j=1}^N \sum_{i=1}^n \mathbb{I}_{\zeta_{i,j}}(g)(\beta'_{j,i} g_i + \beta'_{j,i+n}) =$$
$$\sum_{j=1}^N \sum_{i=1}^n \beta'_{j,i} \mathbb{I}_{\zeta_{i,j}}(g)(g_i - (-\frac{\beta'_{j,i+n}}{\beta'_{j,i}})), = m' \sum_{j=1}^K \sum_{i=1}^L \beta'_{j,i}$$

with $1 \le K \le N$, $1 \le L \le n$. Let us consider a PWP classifier $PWPC(\cdot)$ with parameters $\mathscr{F} = \{f^j\}_{j=1}^N$, such that $f_i^j = -\frac{\beta'_{j,i+n}}{\beta'_{j,i}}$, for all i, j. Then, again, g is classified in the same way by the classifiers $LC_\rho(\cdot)$ and $PWPC(\cdot)$. This is in particular true for $g \in \mathscr{A} \cup T$ and $g \in \mathscr{R} \cup F$. ■